Technical Report

Department of Computer Science
and Engineering
University of Minnesota
4-192 Keller Hall
200 Union Street SE
Minneapolis, MN 55455-0159 USA

TR 12-015

Crime pattern analysis: A spatial frequent pattern mining approach

Shashi Shekhar, Pradeep Mohan, Dev Oliver, and Xun Zhou

May 10, 2012

Report Documentation Page					Form Approved OMB No. 0704-0188

1. REPORT DATE **10 MAY 2012**	2. REPORT TYPE	3. DATES COVERED **00-00-2012 to 00-00-2012**
4. TITLE AND SUBTITLE **Crime pattern analysis: A spatial frequent pattern mining approach**		5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **University of Minnesota,Department of Computer Science and Engineering,200 Union Street SE,Minneapolis,MN,55455-0159**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR’S ACRONYM(S) 11. SPONSOR/MONITOR’S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release; distribution unlimited**		
13. SUPPLEMENTARY NOTES		

14. ABSTRACT

Crime pattern analysis (CPA) is the process of analytical reasoning facilitated by an understanding about the nature of an underlying spatial framework that generates crime. For example, law enforcement agencies may seek to identify regions of sudden increase in crime activity, namely, crime outbreaks. Many analytical tools facilitate this reasoning process by providing support for techniques such as hotspot analysis. However in practice, police departments are desirous of scalable tools for existing techniques and new insights including, interaction between different crime types. Identifying new insights using scalable tools may help reduce the human effort that may be required in CPA. Formally, given a spatial crime dataset and other information familiar to law enforcement agencies, the CPA process identifies interesting, potentially useful and previously unknown crime patterns. For example, analysis of an urban crime dataset might reveal that downtown bars frequently lead to assaults just after bar closing. However, CPA is challenging due to: (a) the large size of crime datasets, and (b) a potentially large collection of interesting crime patterns. This chapter explores, spatial frequent pattern mining (SFPM), which is a spatial data driven approach for CPA and describes SFPM in the context of one type of CPA, outbreak detection. We present a case study to discover interesting, useful and non-trivial crime outbreaks in a dataset from Lincoln, NE. A review of emerging trends and new research needs in CPA methods for outbreak detection is also presented.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**41**	

Crime pattern analysis: A spatial frequent pattern mining approach

Shashi Shekhar, Pradeep Mohan[1], Dev Oliver, Xun Zhou
Department of Computer Science, University of Minnesota, Twin-Cities, MN, USA
{shekhar,mohan,oliver,xun}@cs.umn.edu

Abstract

Crime pattern analysis (CPA) is the process of analytical reasoning facilitated by an understanding about the nature of an underlying spatial framework that generates crime. For example, law enforcement agencies may seek to identify regions of sudden increase in crime activity, namely, crime outbreaks. Many analytical tools facilitate this reasoning process by providing support for techniques such as hotspot analysis. However, in practice, police departments are desirous of scalable tools for existing techniques and new insights including, interaction between different crime types. Identifying new insights using scalable tools may help reduce the human effort that may be required in CPA.

Formally, given a spatial crime dataset and other information familiar to law enforcement agencies, the CPA process identifies interesting, potentially useful and previously unknown crime patterns. For example, analysis of an urban crime dataset might reveal that downtown bars frequently lead to assaults just after bar closing. However, CPA is challenging due to: (a) the large size of crime datasets, and (b) a potentially large collection of interesting crime patterns. This chapter explores, spatial frequent pattern mining (SFPM), which is a spatial data driven approach for CPA and describes SFPM in the context of one type of CPA, outbreak detection. We present a case study to discover interesting, useful and non-trivial crime outbreaks in a dataset from

[1] Corresponding Author

Lincoln, NE. A review of emerging trends and new research needs in CPA methods for outbreak detection is also presented.

1 Introduction

Crime pattern analysis (CPA) is a key step employed by law enforcement and criminal justice agencies towards understanding the spatial environment that generates crime patterns [18]. For example, the analysis of crime datasets with multiple crime types may reveal sudden increase in the activity of a subset of crime types in certain areas. This understanding provides insight into predicting future crime incidents and mitigates existing crimes [8, 37, 38].

The importance of CPA is clearly evident in the growth of spatial crime reports and other spatial information known to law enforcement. Rapid collection and archival of crime reports coupled with the growing analytical needs of law enforcement has given rise to a variety of tools including CrimeStat [43], ArcGIS 10 Spatial Statistics Toolbox [2], GeoDa [6], Rigel [64], SANET [56], SatScan [53] etc.

However, the growing needs of law enforcement stresses scalable ways to generate meaningful crime patterns that may lead to hypotheses regarding the nature of crime as opposed to human driven enumeration of all possible hypotheses. For example, in a typical crime dataset containing 40 different crime types, there may be over 2^{40} different patterns of association between different types. Enumerating all these patterns manually would be an arduous task even for trained analysts. Many police departments aim to accomplish crime mitigation and crime prevention with very few resources. However, the growth in the size and volume of crime datasets poses serious challenges. Hence, there is a growing need for scalable tools that can assist trained analysts and accomplish law enforcement goals with minimal resource allocation. CPA helps law enforcement planners accomplish this goal by identifying interesting, potentially useful, and non-trivial spatial patterns including, regions of sudden increase in crime activity [75,76,77,78,79,81], frequent co-occurrence of crime types around features such as bars [35, 48-50] and crime prone streets [22, 58].

1.1 Existing Research and Our Contributions

Similar to the evolution in scientific methodology [34], CPA has witnessed a massive growth in the number and variety of approaches to identify and interpret patterns of crime [23]. Methodological advancements in CPA include, empirical approaches [31,46,60], criminological theories that interpret the spatial aspects of crime [5,7,10,11,12,21,23,25,26,32,33], offender-profiling techniques to characterize the spatial habitat of offenders [41,44,57,64,65,28,13,14,15,16,17] and crime hotspot detection techniques [3,5,6,27, 40, 42,51,54,56,59,61,62,63,68,69] that can extract concentrations of a single crime type.

A special class of crime patterns is crime outbreak. Crime outbreaks are subsets of a spatial framework with significantly high level of crime activity involving one or more related crime types. The level of significance in a crime outbreak is calculated based on a ratio between observed count within a region and the count expected via a user specified probability model [53,61,76,77,78,79] as opposed to simple counts or local autocorrelation statistics that are used in traditional hotspot detection methods. Outbreak detection has been a explored in Spatial epidemiology via statistical models to compute the likelihood of different disease types in being a part of any disease outbreak [75]. Figure 1 shows a classification of different CPA approaches to outbreak detection. As shown in this figure, CPA methods for outbreak detection can be classified into two broad categories. The left category corresponds to techniques that are designed to analyze crime outbreaks in a single crime type or entire crime datasets without any type level stratification. These techniques include, hotspot analysis methods for detecting crime hotspots of points [40,42,43,52] areas or zones [5,6,54] and street network subsets [7,22,56,58]. The rationale for this classification of crime hotspots is based on the type of hotspot map that is produced [27]. However, in most crime analysis scenarios, an understanding of the co-occurrence of different crime types and spatial features such as bars may reveal interesting patterns pertaining to geographical locality of certain crime types [27]. Existing techniques in hotspot analysis are not designed to handle this scenario.

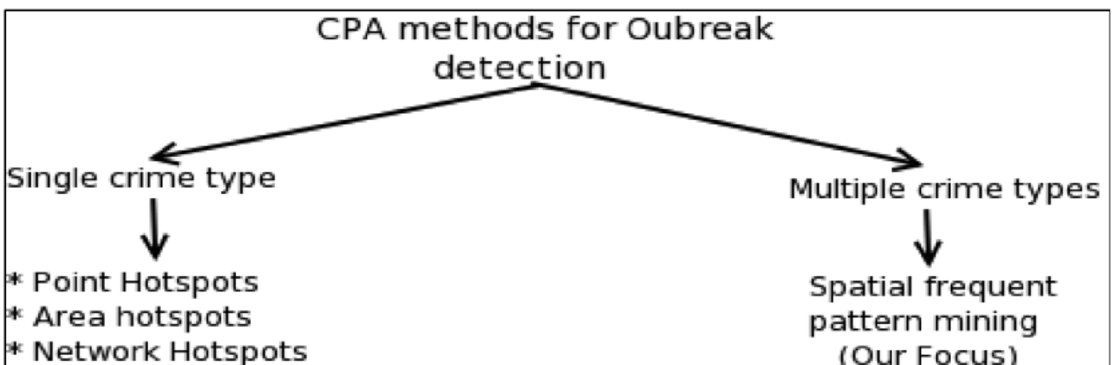


Figure 1: Classification of CPA methods for Outbreak detection

In contrast, the proposed spatial frequent pattern mining (SFPM) is designed for analyzing patterns involving multiple crime types and their associations with different features such as bars. In this chapter, we describe one class of SFPM methods, namely, regionally frequent crime patterns (RFCP) that are spatial representations of crime outbreaks involving multiple crime types. Formally, RFCPs are subsets of different crime types that co-occur frequently in certain areas of a spatial framework. In order to quantify the significance of a multi-type crime outbreak, we make use of the Multinomial scan statistic [75]. The multinomial scan statistic extracts significant crime types that may be involved in one or more outbreaks via a maximum likelihood ratio test statistic. Based on the significant crime types identified, the RFCP discovery process identifies spatial subsets of these significant crime types and quantifies crime outbreaks using a regional conditional probability (RCP) measure.

This chapter makes the following contributions:

(a) We define spatial frequent pattern mining (SFPM) and illustrate it using one class of methods namely, regionally frequent crime patterns (RFCPs).

(b) We describe the methodology for discovering crime outbreaks represented as RFCPs by making use of the multinomial scan statistic. We illustrate this methodology with a case study using a crime dataset from Lincoln, Nebraska and present discovered crime outbreaks involving multiple crime types.

(c) We describe emerging trends in spatio-temporal frequent pattern mining in the context of space-time outbreak detection.

(d) We outline research needs including, methodological advancements to account for special semantics of spatio-temporal data and key aspects urban crime scenarios, including analyzing crime distributions along street networks.

1.2 Scope and Organization

This chapter provides an overview of spatial frequent pattern mining and outlines its capabilities for aiding in law enforcement with the primary intent of reducing human effort that may be required in crime pattern analysis. Hence, this chapter does not focus on presenting computational performance evaluation. There are a number of crime pattern analysis methods in criminology. However, this chapter just focuses on outbreak detection in crime data.

The chapter is organized as follows: (a) Section 2 reviews some basic concepts including, defining spatial frequent pattern mining and describes regionally frequent crime pattern (RFCP) with examples. (b) Section 3 describes our methodology to discover crime outbreaks via RFCP mining. (c) Section 4 describes some emerging trends in spatio-temporal frequent pattern mining, and (d) Section 5 outlines some research needs including, new methodological advancements in crime pattern analysis for outbreak detection.

2 Basic Concepts

This section reviews some concepts related to spatial frequent pattern mining (SFPM) and defines the crime outbreak detection problem.

Spatial frequent pattern mining (SFPM) is the process of discovering interesting, useful and non-trivial patterns from spatial datasets. Figure 2 shows a typical SFPM process that is based on the crime datasets collected and archived by law enforcement as the basis. The SFPM process usually begins with knowledge of criminological theories from environmental criminology. Based on these theories, analysts pose certain questions on the data. Some questions would be: *Why do downtown bars often lead to assaults crimes but bars in other regions seldom do so? Is it regional differences in geographic concentration? Are there regional differences in patron demography or crowd density? Are there policy differences in screening, bouncing, policing?*

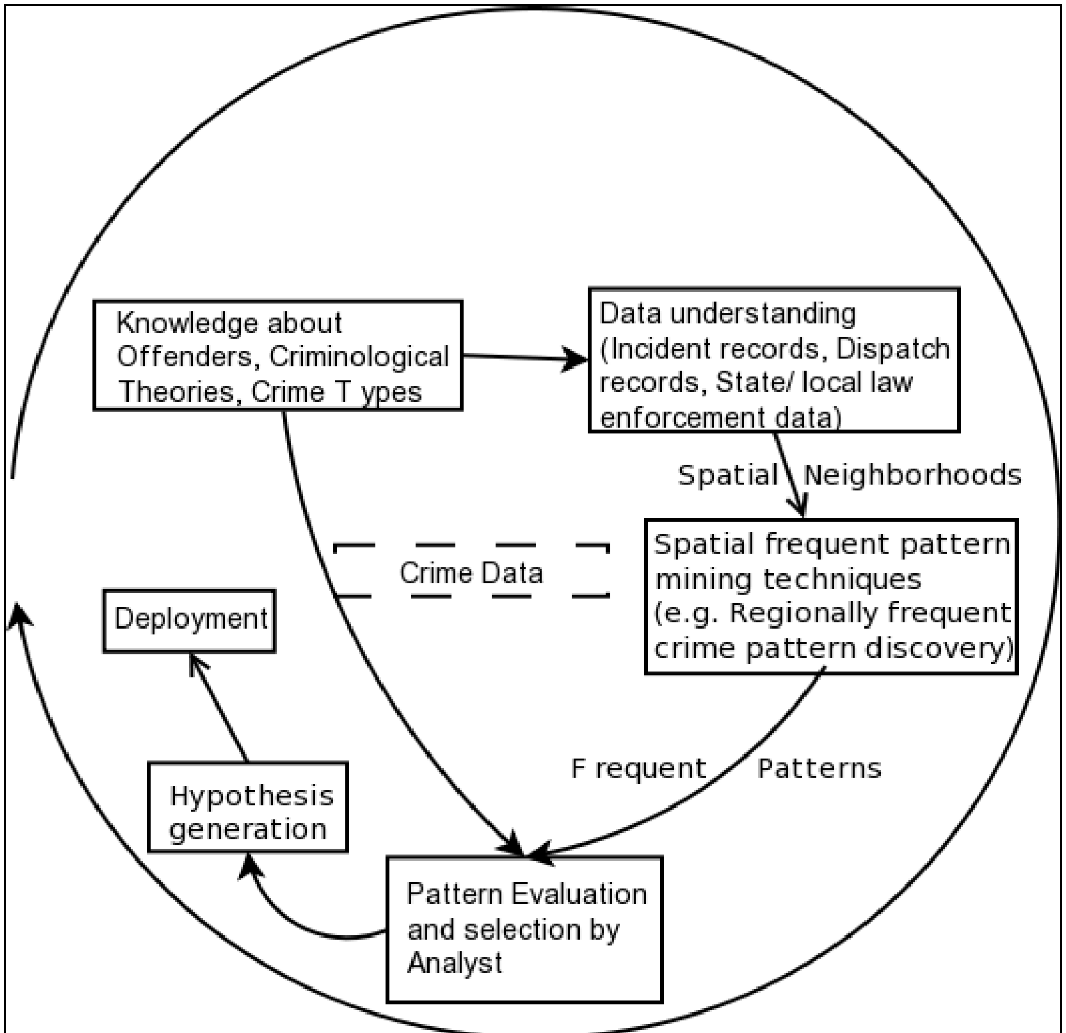
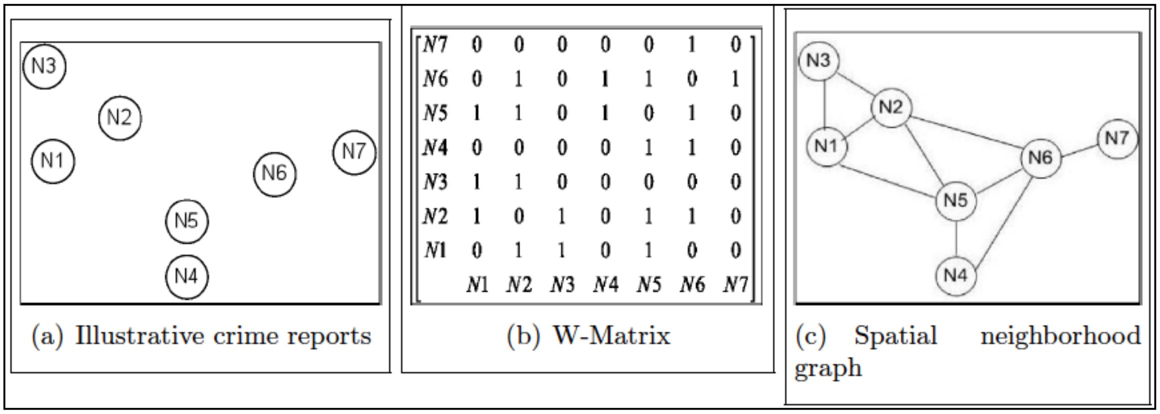


Figure 2: Spatial frequent pattern mining process

Based on these questions, the SFPM process involves understanding the available data via suitable interpretation models, including fitting of well known distributions, specifying suitable representations of observed crime patterns, understanding of possible neighborhood effects. A key outcome of the data understanding step in SFPM is the knowledge of spatial neighborhoods that may be useful to explore possible relationships between individual data entities such as crime reports, police districts, crime prone streets etc. A **spatial neighborhood** is a collection of related spatial entities such as crime reports. A spatial neighborhood is usually specified by means of a **spatial neighbor**

relation. A spatial neighbor relation can be topological or distance based [66].



(a) Illustrative crime reports (b) W-Matrix (c) Spatial neighborhood graph

Figure 3: Spatial neighborhood matrix and spatial neighborhood graph

In most crime analysis applications involving crime reports, the most common type of neighbor relation is a distance-based relation. The application of a spatial neighbor relation on a collection of crime reports produces **a spatial neighborhood matrix**, commonly referred to as the **W-Matrix**. For example, Figure 3(a) shows an illustrative crime report dataset showing different crime reports represented as circles with labels **N**1, **N**2 etc. Application of a neighbor relation based on a distance threshold (e.g., 1 mile, 2 mile etc.) produces a spatial neighborhood matrix as shown in Figure 3(b). The matrix in this figure consists of 0s or 1s to represent the absence or presence of a neighbor relationship. Figure 3(c) shows an alternative representation of the W-Matrix called **the neighborhood graph**, where the edges represent the presence of a neighbor relation and the nodes represent the crime reports.

Given spatial crime data, neighbor relationships, environmental criminology theories and other inputs known to law enforcement, SFPM employs several techniques to identify interesting, useful and non-trivial crime patterns. One such technique is

Regionally frequent crime pattern (RFCP) discovery.

RFCPs represent collections of spatial features and crime types frequently associated with each other at certain localities. For example, the RFCP, <*(Bar, Assaults), Downtown* > indicates that a frequent pattern involving assaults and bars is often localized in downtown regions. Given feature types[2] (e.g., Bars), crime types and their geo-located instances, along with a spatial neighborhood size and a likelihood threshold, the RFCP discovery process finds all interesting RFCPs. For example, Figure 4(a) shows an illustrative crime dataset consisting of one feature type, Bars, and two crime types, Assaults and Drunk Driving. Red circles represent bars; blue triangles represent assaults, green squares represent drunk driving.

Given the input in Figure 4(a), the RFCP discovery process identifies RFCPs as shown in Figure 4(b). In the Figure, the RFCPs (shown) correspond to the collection {ABC}, which consists of crime types, bars, and their localities.

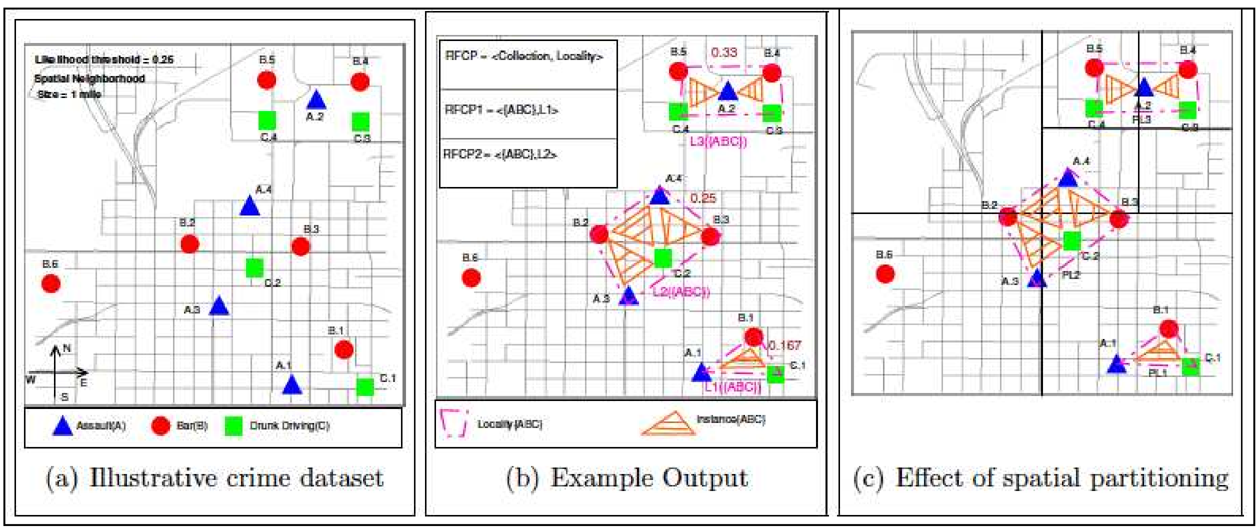


Figure 4: Illustrative example of RFCP (Best viewed in color)

[2] For brevity, the term “feature types" and “feature" are used interchangeably in this paper. For example, a bar feature may correspond to a bar feature type such as bar closing or happy hour, etc., that occurs at a bar location

This RFCP is represented as <{ABC}, Locality> in Figure 4(b). The locality is shown as dotted purple polygons. The dotted purple polygons represent localities where the collection of crime types and spatial features, {ABC} may be frequent. The shaded orange triangles represent individual instances of the collection **{**ABC**}**. In Figure 4(b) the localities are defined based on the convex hull of related crime reports and spatial features. One may also choose to specify these localities by partitioning the spatial framework via quad partitions as shown in Figure 4(c). However, such a scheme may result in loss of information about spatial relationships at the boundary of such a partition.

The number adjoining each locality in Figure 4(b) is the chance that, the collection **{**ABC**}** containing instances of Bar feature can be observed along with instances of crimes like assault or drunk driving. This chance of occurrence can be interpreted as a *local fraction* of instances corresponding to a crime type or feature type that *participates in a collection* [35, 50]. A collection such as **{**ABC**}** along with its observed locality may be considered an RFCP if the chance of it occurring is above the user specified threshold. The local fraction of instances of any crime type participating an RFCP is measured using a **Regional conditional probability (RCP).** RCP can be defined as follows:

$$\text{RCP(Crime-type,RFCP)} = \frac{\#\,\text{instances of Crime-type participating in RFCP}}{\#\ \text{instances of Crime-type in the Dataset}} \qquad (1)$$

Since, an RFCP is a collection of co-occurring crime types in the vicinity of spatial features like bars, quantifying the importance of an RFCP involves accounting for the location fraction of different crime types. In typical crime datasets, 80% of the crime might occur in 20% of the places. Hence, one may be interested in particular bars that may be responsible for a large number of crimes. Hence, the actual measure that quantifies the interestingness of an RFCP is the **lower bound on the regional conditional probabilities** of all crime types and features that *participate* within a pattern. This lower bound can be termed as the **Regional conditional probability index (RCPI).** This measure can be defined as follows:

$$\text{RCPI(RFCP)} = \min\left\{\frac{\#\text{instances of Crime-type participating in RFCP}}{\#\text{instances of Crime-type in Dataset}}\right\} \qquad (2)$$

The **min** term in (1) represents that the value of RCPI is the least RCP of all crime types or features within a RFCP. For example, for the RFCP in Figure 4(b), <{ABC}, L2> contains two instances of Assault represented by the letter A. Since there are 4 instances of Assault in the dataset shown in Figure 4(a), the RCP of A within the RFCP <{ABC}, L2> is 2/4. That is the local fraction of Assault within the RFCP is 2/4, which is 0.5. This implies that the chance of witnessing an Assault crime within the locality L2 is 50%. Similarly, the RCPs for Bar feature and Drunk Driving arrests represented by the letter C is 2/6 and 1/4 respectively. Hence, the RCPI of the RFCP, <{ABC}, L2> can be computed as follows:

$$\text{RCPI}(<\{\text{ABC}\},\text{L2}>) = \min\left\{\frac{2}{4},\frac{3}{6},\frac{1}{4}\right\} = \frac{1}{4} = 0.25$$

The value 0.25 is the degree of interestingness of the RFCP <{ABC},L2} and represents the chance that the entire RFCP may occur when at least one instance of any of the participating crime types or spatial features occur within locality L2.

Based on the definition RFCPs and their interestingness, a **crime outbreak** of multiple crime types may be viewed as an RFCP, which has participation from crime types that have a high chance of participating in a **significant cluster** within the spatial framework. Spatial statistics has explored a popular statistical, the **multinomial scan statistic** [75,81] that can identify crime types that have a high chance of participating in significant clusters [75].

Based on the above definitions, we define and illustrate crime outbreak detection as an analysis problem that may require a solution using SFPM .

2.1 Crime outbreak detection and Illustration

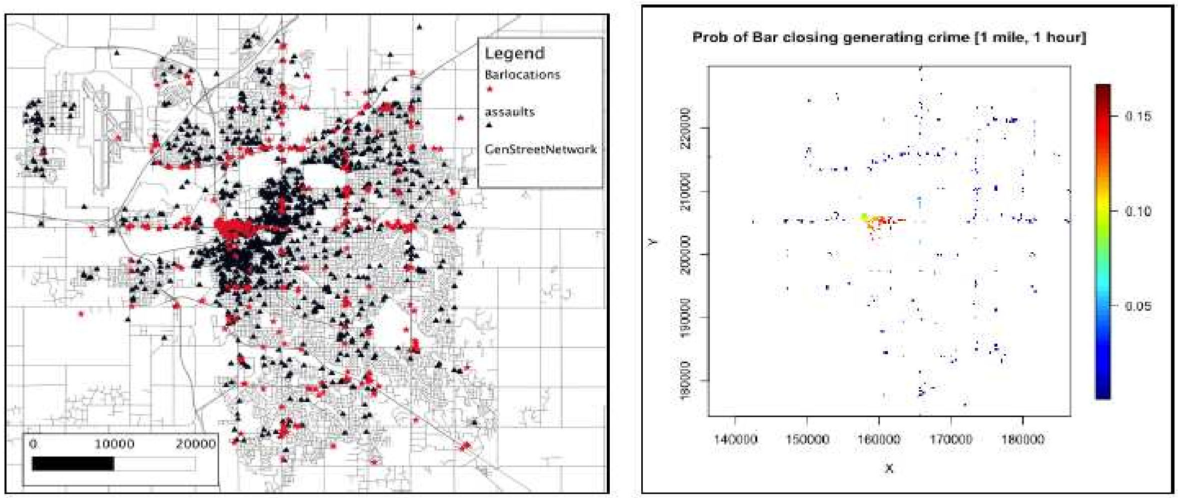
In this section, we define crime outbreak detection as a data analysis problem and illustrate it with an example based on a crime dataset from Lincoln, Nebraska [1]. Formally, crime outbreak detection requires the following inputs specified by an analyst:

- **(a)** A spatial crime dataset with a collection of crime types and others spatial features including, bars, schools etc.
- **(b)** Other ancillary information known to law enforcement, spatial neighborhoods and appropriate interestingness threshold.

Based on the above inputs, the goal of crime outbreak detection is to report a collection of RFCPs with RCPI values greater than interestingness threshold specified by the analyst. However, an important constraint is that, the identified RFCPs contain crime types and spatial features are highly likely to be a part of a significant crime outbreak.

2.1.1 Problem Illustration

A typical input to crime outbreak detection is a spatial crime dataset that may contain two or more crime types and many spatial features such as bars, parks, schools etc. For example, Figure 5(a) shows a crime dataset from Lincoln, Nebraska.[1] This Figure shows bar locations as red circles and assault crime report locations as black triangles. The Lincoln crime dataset contains about 40 different aggregate crime types with an average of 5-6 sub-types for each crime type. The entire dataset corresponds to the years 1999-2007.The assault crime reports shown in Figure 5(a) are drawn from the year 2007. Figure 5(b) shows the count of assaults within a 1-mile vicinity of bars within



(a) Bar locations and Assault incidents in 2007 (b) Probability of Bars generating crime after bar closing

Figure 5: Subset of Crime data from Lincoln, NE, USA, 2007[1](Best viewed in color)

an hour after bar closing. This figure reveals that bars in the center (downtown Lincoln) have a high chance of assaults possibly representing an outbreak. Apart from crime reports, spatial features and spatial neighborhoods, analysts are also required to specify thresholds of interestingness to discover RFCPs(e.g. 0.25, 0.15, 0.001 etc.).

Based on the above inputs, the proposed SFPM technique (i.e. RFCP discovery) identifies regions where different crime types co-occur in the vicinity of bars. For example, Figure 6(a-d) shows a typical output of the analysis. This figure shows RFCPs corresponding to two crime types, Assaults and Larceny and one feature type Bar. The RFCPs containing different subsets of types are shown as blue polygons.

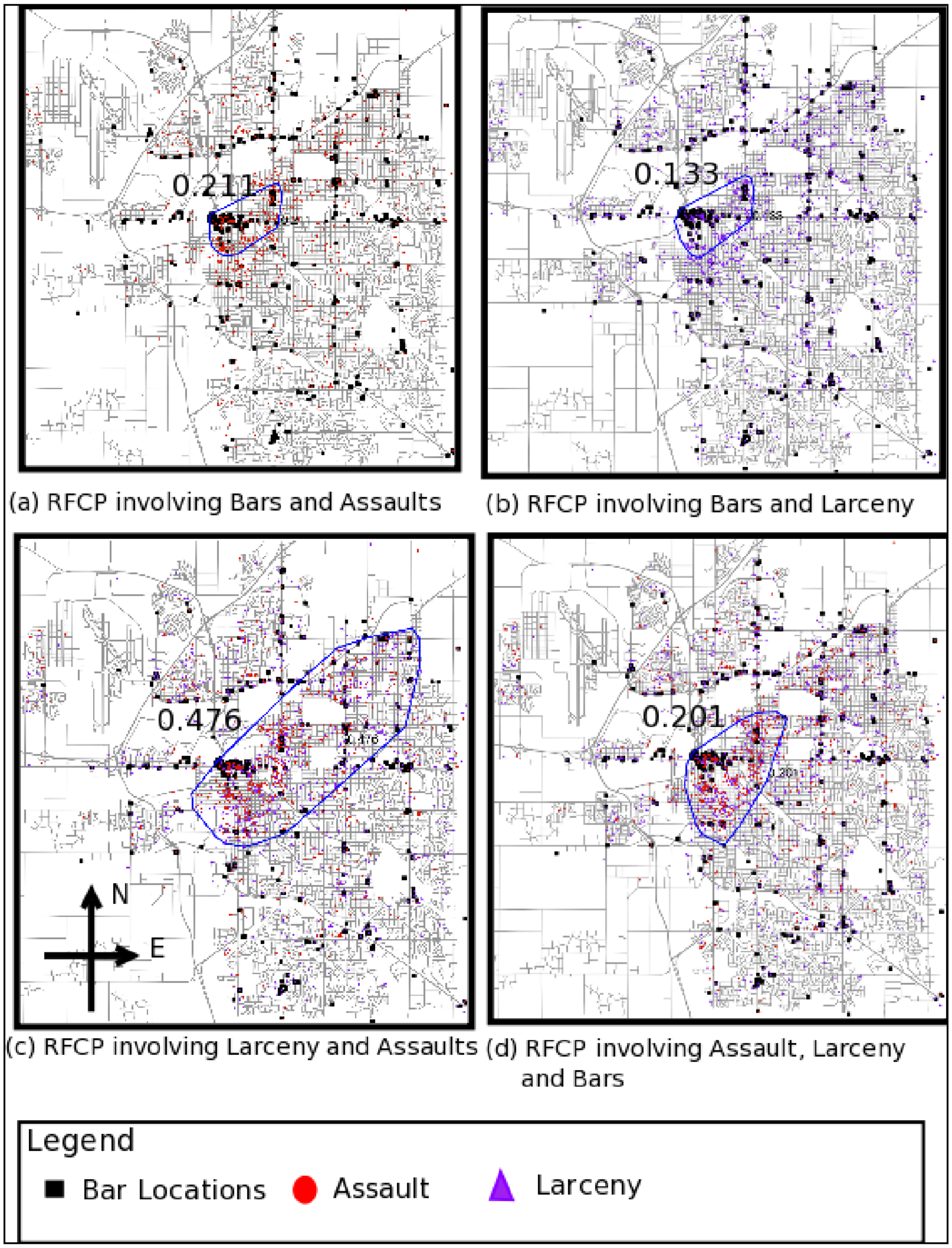


Figure 6: Typical output of crime outbreak detection via RFCP discovery

Figure 6(a) shows that the chance of witnessing Assaults in the vicinity of downtown bars is 0.211 or 21.1 %. Figure 6(b) shows that the chance of witnessing Larceny crimes in the vicinity of Downtown bars is 13.3%. On the other hand, Figure

6(c) shows that the probability of finding Larceny crimes around the site of an assault is higher with a chance of 47.6% and includes both the downtown area and the northeast area of Lincoln, NE. Figure 6(d) shows an RFCP corresponding to Bars, Assaults and Larceny. This figure shows that the chance of finding related assaults and larceny crimes in the vicinity of bars in downtown is increased to a chance of 0.201 or 20.1% with respect to the one shown in Figure 6(b). For example, Figure 6(b) reports a lower chance of finding a larceny in the vicinity of bars downtown.

An important in constraint in Crime outbreak detection via RFCP discovery is that, the crime types participating in RFCPs need to have a significantly high chance of being a part any outbreak within an RFCP.

3 Crime outbreak detection: Proposed Approach

In this section, we present an overview of crime outbreak detection using RFCP discovery and illustrate it via a case study from the Lincoln, crime dataset.

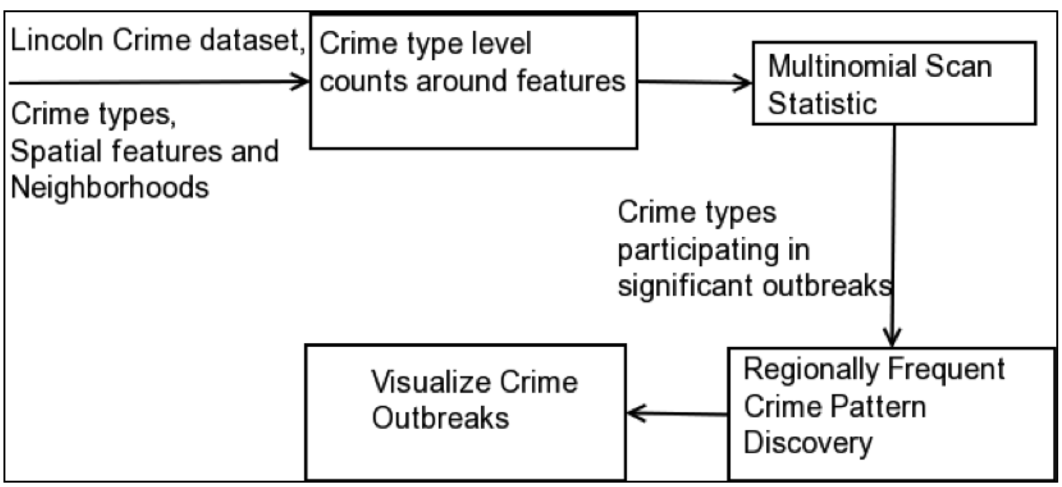


Figure 7: Overview of proposed approach

Figure 7 shows an overview of the proposed approach to detect crime outbreaks. The Lincoln crime dataset from the year 2007 contains crime reports with multiple crime types and several feature types such as bars. In addition, the user also provides a sense of spatial neighborhood (in terms of the maximum neighborhood size, typically 0.5 miles, 1 mile etc.). Based on these inputs, the first stage, computes crime type level counts around each spatial feature.

Once these counts are obtained, we make use of the Multinomial scan statistic test implemented in the SatScan program [53]. The multinomial scan statistic routine in SatScan computes several interesting measures including, (a) spatial features that may be a part of significant outbreaks, (b) Radius of a possible crime outbreak cluster around each feature and (c) The risk of occurrence for each crime type within each outbreak cluster. The risk value for each crime type corresponds to the ratio of, the number of cases the crime type within the radius of the feature to the number of cases of the crime type expected to occur based on a standard multinomial distribution. In our analysis we consider only crime types that have a high risk of occurring within the neighborhood of the spatial feature(i.e. risk > 1). A detailed description of the notion of risk and its interpretation can be found in the SatScan manual [53]. To ensure that only crime types participating in statistically significant outbreak are highlighted, SatScan performs Monte Carlo simulation and computes p-values. Assuming a standard significance level of 0.05, only significant spatial features and crime types that have a high-risk level within their neighborhood are retained after the multinomial scan statistic test.

The results of the multinomial scan statistic test include, crime types that may be involved in one or more interesting crime outbreaks. However, it is still important to extract the actual regions of these outbreaks. This goal accomplished via RFCP discovery. The RFCP discovery process requires other inputs including interestingness thresholds, and spatial neighborhood information. Based on this the RFCP process reports all interesting RFCPs of crime types that may participate in one or more outbreaks. Since RFCPs also include the actual location of the crime report, they provide an enhanced spatial view (e.g. convex polygon) of the actual outbreak as opposed to a simple circular neighborhood.

3.1 Analysis Results on Lincoln Crime dataset

The crime dataset from Lincoln, NE in the year 2007 has over 38000 crime reports and 40 crime types. For the purpose of this analysis, bars in Lincoln were the only spatial features included. Lincoln city has 403 bars. To determine type level counts in the vicinity of each bar, we used a spatial neighborhood based on a distance threshold of 0.5 miles. This step is similar to performing an overlay operation within a GIS. However, due to a large number of crime types, the overlay operation needs to be performed using a specialized spatial database [66] with indices geometric built on the geometry of different crime reports. A manual task of performing overlay within a GIS would require enormous human effort, while the use of a spatial database reduces the effort and time required to compute type level counts in the vicinity of each bar location.

Type level counts at each bar location and the co-ordinates of the bars were then provided as input to the multinomial scan statistic routine in the SatScan program [53]. The multinomial scan statistic requires a maximum spatial neighborhood search radius. We used the same search radius that was used to compute the type level counts, i.e. 0.5

miles. Based on these inputs SatScan computes the multinomial scan statistic for each bar location and performs montecarlo simulation to generate p-values. Only bars that are a part of significant crime outbreak (i.e. p-value <=0.05) were chosen for further analysis. Also, for each bar location, SatScan, computes the relative risk level of a crime type belonging to that outbreak. Each crime type in the dataset may belong to one or more outbreaks across Lincoln. The number of outbreaks for each crime type is shown in Figure 8. SatScan also reports a radius around each bar location to provide a spatial

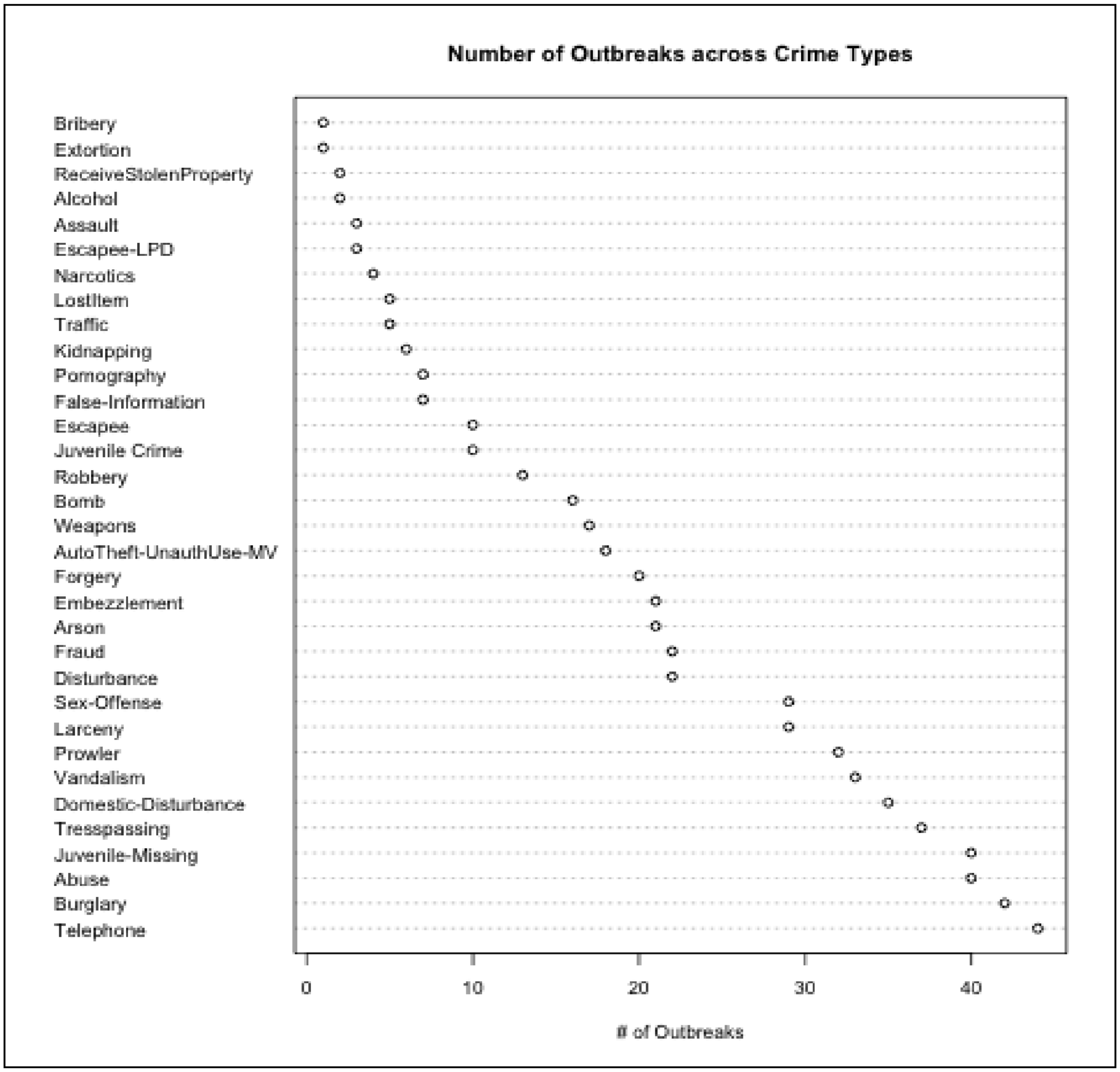
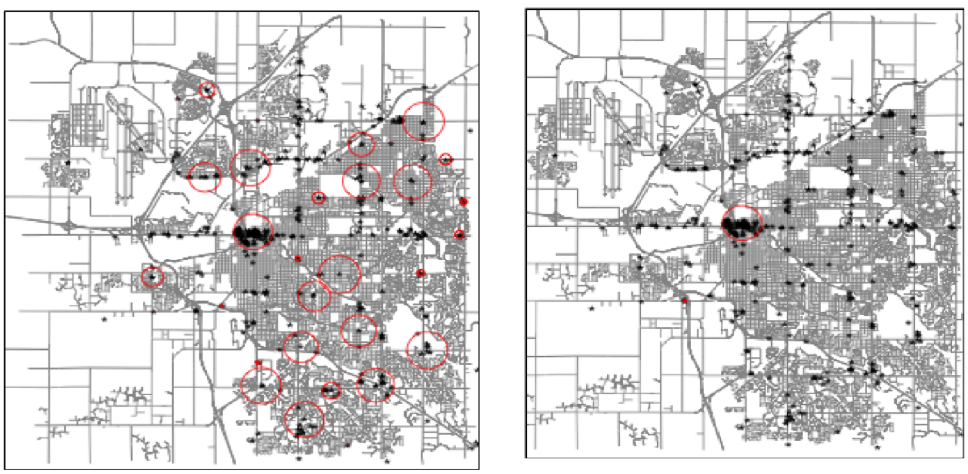


Figure 8: Number of outbreaks for each crime type in Lincoln, NE, 2007.

extent for the possible crime outbreak. Based on this information, one can visualize the extent and location of crime outbreaks for each crime type. Figure 9 shows the location and extent of possible outbreaks corresponding to two crime types, namely, vandalism and alcohol crimes (spatial outbreaks of other crime types are shown in Appendix I).In Figure 9, the black asterisks represent the location of bars and the red circles represent possible crime outbreaks. The radius of the circles is determined by multinomial scan statistics.



(a): Vandalism outbreaks **(b): Alcohol outbreaks**

Figure 9: Location and extent of crime outbreaks by crime type, Lincoln, NE, 2007

However, the circles shown in Figure 9, provide only a rough estimate of the outbreak because they do not take into the actual location of different crime types. A visual examination of Figures 9(a) and 9(a) reveals that vandalism crime outbreak (Figure 9(a)) and alcohol crime outbreaks (Figure 9(b)) possibly co-occur at the center of Lincoln (the cluster of bars at the center). To identify the final crime outbreak regions, we extract all crime types from the dataset that have a high risk of occurrence in or more outbreaks (Due to space limitations, the risk levels for different crime types is illustrated in Figure

A2 in the Appendix). Crime types that may be a part of significant outbreaks can be considered as candidates for input to the RFCP discovery process.

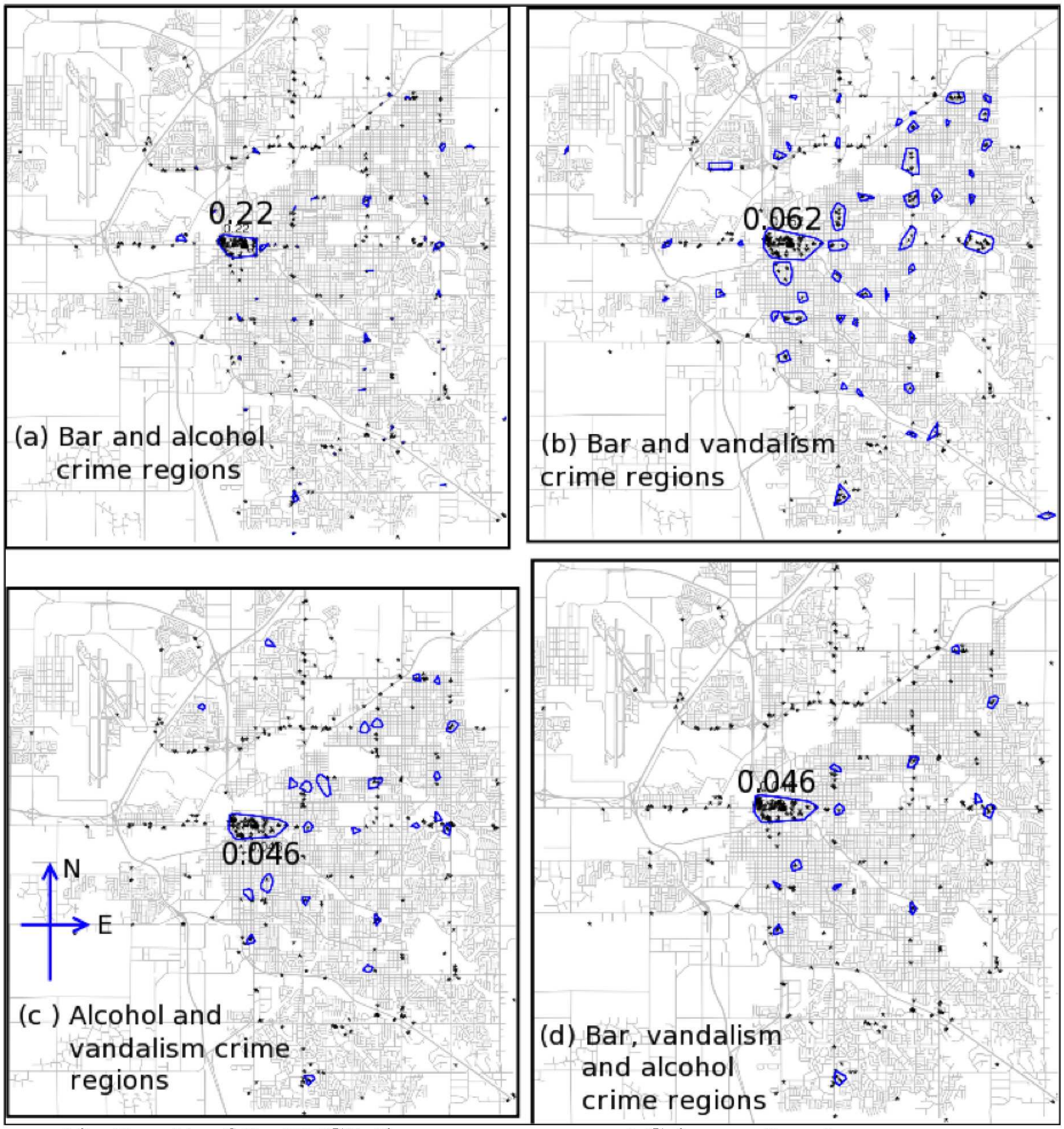


Figure 10: Results of the RFCP discovery process and Crime outbreaks

In addition to the crime types involved in significant outbreaks, the RFCP discovery process requires additional such as the spatial neighbor relation and an interestingness threshold. For discovering RFCPs , we used a neighborhood distance of

700 feet and an interestingess threshold of 0.001. The results of the RFCP discovery process is shown inf Figure 10. Figure 10(a) shows the RFCP corresponding to Bars and Aclcohol crimes. The bars that were highlighted in Figure 9(b) appear again in this Figure indicating that, the downtown bars may be responsible for all the outbreaks in alcohol crime. This implies that, bars in Lincoln have a 22% probability (i.e. a value of 0.22) of leading to an alcohol outbreak and most of these bars are localized in dowtown Lincoln. Similarly, Figure 10(b) shows the RFCPs corresponding to Bars and Vandalism crimes with the downtown bars showing highest chance of leading (i.e. 0.062 or 6.2 %) to a Vandalism outbreak. Similarly, Figures 10(b) and 10(c) show a possible association between Vandalism and Alcohol crime outbreaks particularly in the vicinity of downtown bars. While, the multinomial scan statistic's results illustrated in Figure 9 show an indirect association between Vandalism and Alcohol, the RFCP analysis directly reveals a possible association indicates that the highest chance of this association is localized in dowtown Lincoln.

4 Emerging Trends: Space-time crime outbreaks

Crime outbreaks can also happen over space and time. Crime pattern analysis literature has explored space-time hotspots via conceptual constructs such as the hotspot matrix and provides several useful suggestions to police for designing effective mitigation plans [62]. A space-time outbreak is a similar notion but may be defined over multiple crime types. The SatScan program's multinomial scan statistic allows for detecting crime types that may be a part of outbreaks over space and time. To identify such outbreaks, SatScan requires a temporal search window for detecting space-time outbreaks. Based on these inputs SatScan identifies bar locations and days at which these locations may have crime outbreaks.

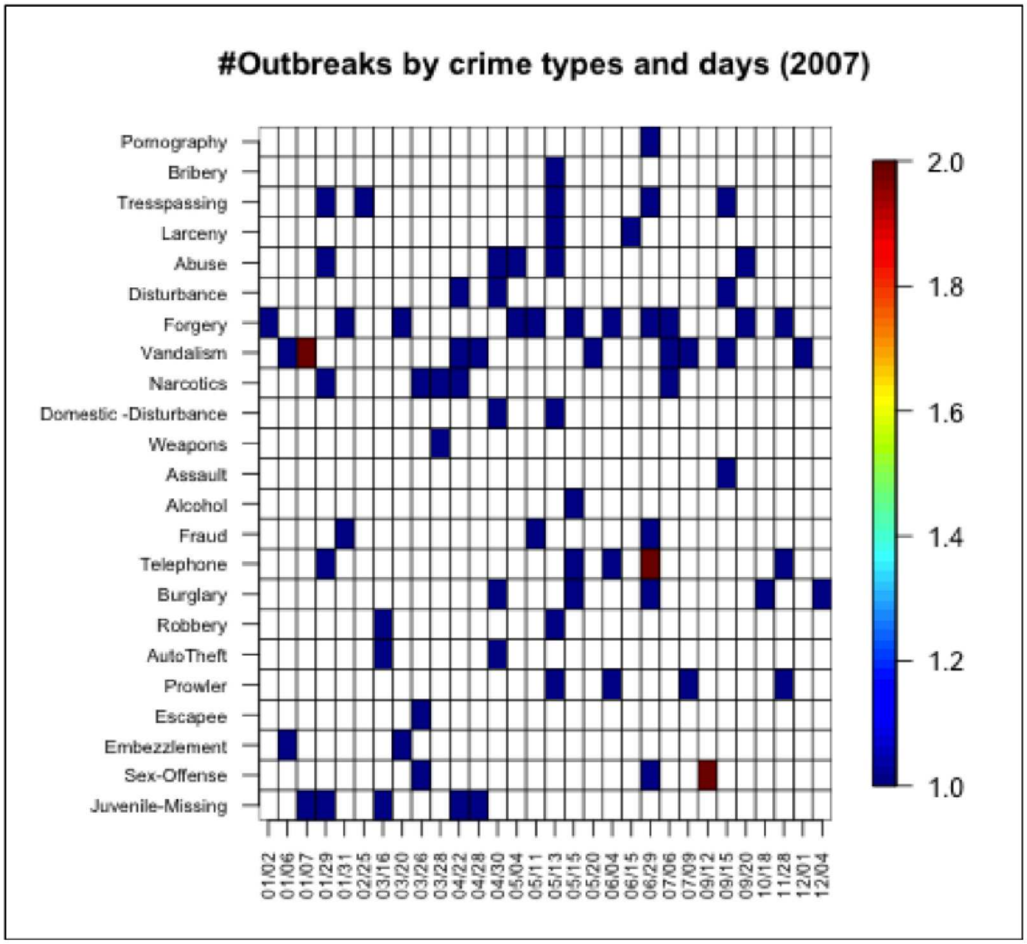


Figure 11: Results of space-time multinomial scan statistics (Best viewed in Color)

Figure 11 shows the result of space-time multinomial scan statistics on the Lincoln crime dataset. The X-axis in Figure 11 shows different days in the year 2007. These are days that may have one or more crime outbreaks as opposed other ordinary days in the year 2007. The SatScan program identified these days based on a temporal search window input of one day. The Y-axis corresponds to crime types that may participate in one or more outbreaks on the days shown in the X-axis. The color in Figure 11 represents the number of outbreaks corresponding to different crime types. For example, there were 2 outbreaks of the vandalism crime on January 7th 2007. The risks of each crime type participating in an outbreak can also be computed (shown in appendix Figure A.3).

The days shown in the X-axis of Figure 11 may correspond to important events including, conventions, football games etc. For example, the day, 15th September (09/15) corresponds to a Nebraska Cornhuskers football game [45] and shows outbreaks in 4 crime types. Also, when datasets with finer temporal granularity (e.g. time stamps) are considered, routine events such as bar closing may also lead to sudden increase in crime. However, in order to account for patterns in such fine temporal granularity, one may have to have to consider topologically rich space-time pattern semantics. Figure 12 shows one such spatio-temporal (ST) pattern, Cascading ST patterns [48].

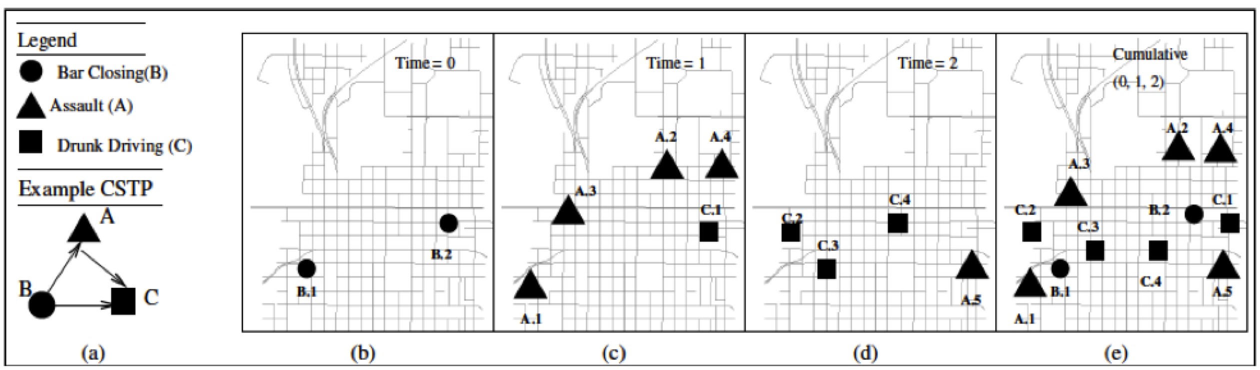
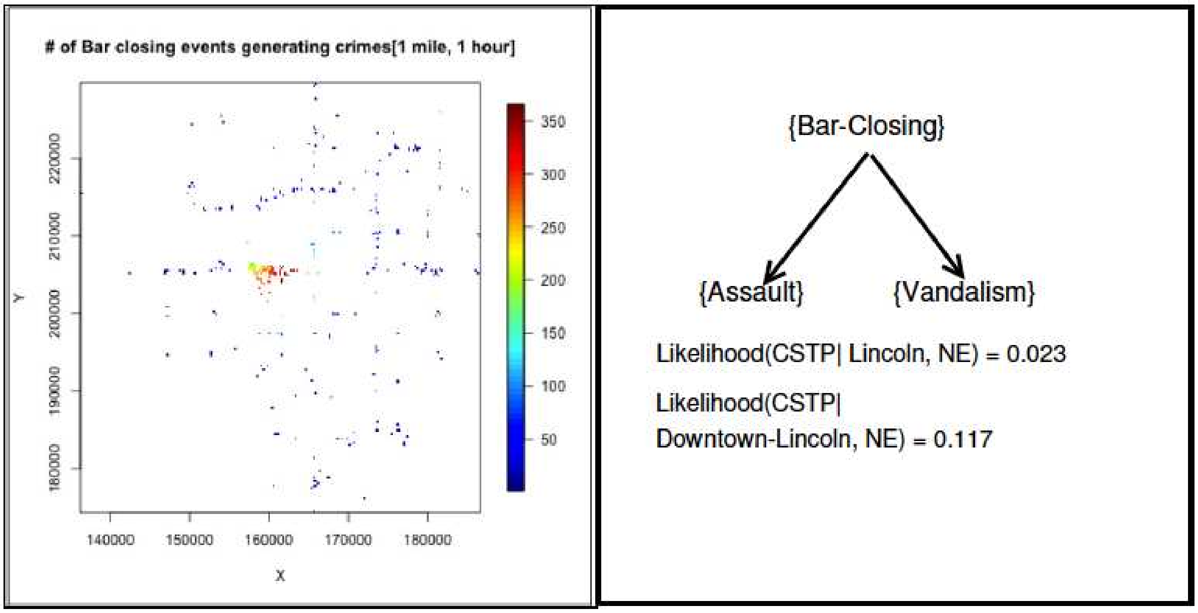


Figure 12: Illustrative ST crime dataset and cascading ST pattern

CSTPs represent partially ordered (over time) collections of spatio-temporal event types that are frequently associated with each other in a locality and occur in a series of stages over time. Figure 12(a) shows an example of a CSTP from an urban crime dataset. The first event here is bar-closings (represented by circles), and subsequent events are assaults (represented by triangles) and drunk driving (represented by squares). Figures 12 (b), (c) and (d) show individual instances of events of bar-closing, assault and drunk driving with their location and time. Figure 12(e) shows the locations of all event instances. When applied to the Lincoln crime dataset, CSTP discovery may be able to



(a) Frequency of Crimes after bar closing

(b) CSTP discovered from Lincoln, NE

Figure 13: A CSTP from Lincoln, NE , 2007 Dataset (Best viewed in color)

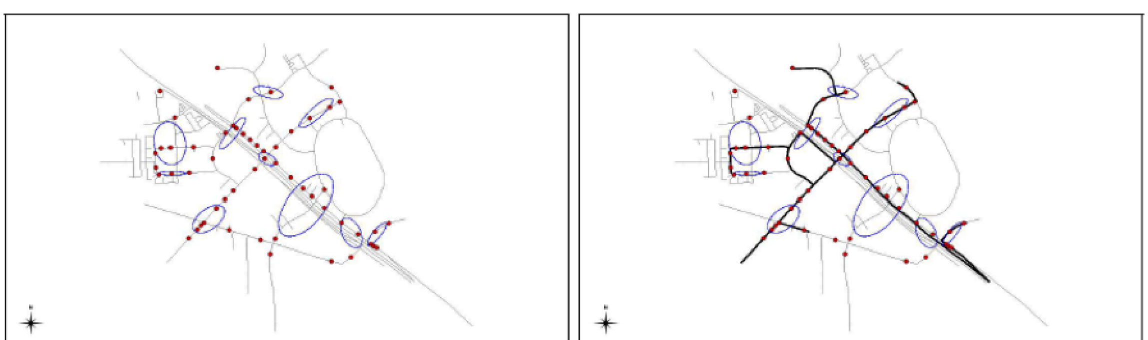
reveal bar closing events that frequently lead to crimes including, assaults and vandalism. Discovering CSTPs may help crime analysts generate hypothesis about crime generators and crime attractors [70]. Using CSTPs, law enforcement can possibly find a related series of crimes and plan intervention strategies during a large crowd gathering (e.g., football games). In addition to identifying activation times of potential crime generators or attractors, CSTPs also help understand appropriate times at which specific crime types may happen around such sources. For example, Figure 13 shows one such CSTP discovered from Lincoln, NE. Figure 13(a) shows a map of bars in Lincoln, NE with the color representing the number of bar closing events that lead to an assault or vandalism. Figure 13(b) shows the CSTP corresponding to the map in Figure 13(a) and lists the chance of finding this pattern in entire Lincoln city and only within downtown. Not surprisingly, finding this pattern is more likely in downtown Lincoln rather than in other areas due to the large number of bars in the center.

5 Research Needs

In this section, we review some important research needs that include methodological advancements that may be needed to effectively analyze crime patterns. These include, techniques for the detection of crime outbreaks along street networks and outbreak detection at multiple analysis scales.

5.1 Crime outbreaks along street networks

Most existing methods for crime outbreak detection involving multiple crime types are largely based on Euclidean spatial neighborhoods. However, for most practical situations it may be interesting study the outbreak of crimes along a street network. Detecting such outbreaks may provide insights to law enforcement analysts to identify possible patrolling districts after taking a variety of other factors into consideration. Environmental criminology uses concepts such as Nodes, paths and edges [11, 23] that may be frequently used by offenders in their journey to crime. Identifying frequent crime routes taken by a vast majority of offenders helps law enforcement make planned changes to their monitoring patterns.



(a) Output of CrimeStat K-Means [43]

(b) Overlayed output comparison with KMPR

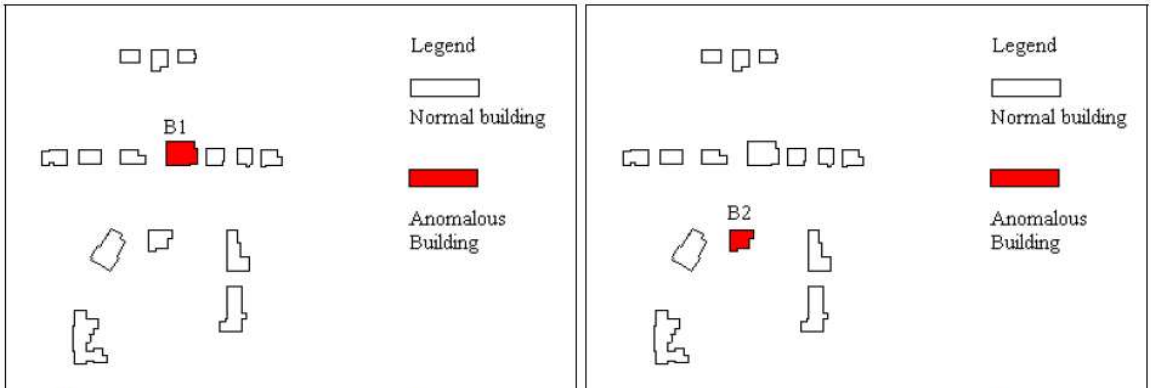
Figure 14: Route level summarization of crime outbreaks(Best viewed in color)

Hence, a route level summarization of crime outbreaks may be more intuitive compared to a geometric (e.g. ellipsoidal or circular) summarization of outbreaks. Figure 14 shows the result of this summarization. It is very clear from Figure 14(b) that a network level summarization may be more intuitive as opposed to an ellipsoidal scheme used tools such as CrimeStat. In this figure the ellipsoidal scheme ignores the network topology and clusters crime reports across the network. Using the result of route summarization in Figure 14 and a variety of other factors, analysts may be able identify appropriate patrolling routes to mitigate crime.

One can hand map these patrol routes by taking a variety of relevant factors into consideration. However, this might require enormous human effort. Instead, one can explore solutions based on SFPM to reduce the possible routes and pick the most suitable patrol districts with aid of an analyst. Given a collection of crime reports, a street network and a user specified parameter (K), SFPM has explored techniques for summarizing crime reports along a street network [56, 58]. However, these techniques may have to be revisited to account for multiple crime types that may participate in significant crime outbreaks.

5.2 Outbreak detection at multiple analysis scales

Handling spatial scale has been an open research challenge in many GIScience applications [73]. In crime outbreak detection one of they key inputs specified by analysts is the spatial neighborhood size. The results of the any analysis are sensitive to the neighborhood size specified by the user. Particularly, in the approach shown in Figure 7, the crime type level counts are sensitive to the spatial neighborhood size. Also, the RFCP process requires a spatial neighborhood size as an input. This makes any spatial analysis technique sensitive to spatial scale. For example, Figure 15 shows a simple scenario for detecting outlier buildings from a subset of the Lincoln, NE dataset using their area attribute.



(a) Map of outlier building $B1$ (highlighted) and its neighborhood

(b) Map of outlier building $B2$ (highlighted) and its neighborhood

Figure 15: Map of two outlier buildings in Lincoln crime dataset (best viewed in color)

Figure 15(a) and (b) are the analysis performed at different spatial scales, namely, two nearest neighbors and eight nearest neighbors respectively. Outlier analysis using the first scale, two nearest neighbors, highlights the building ***B1*** as anomalous (Figure 15(a)), whereas, when we increase the scale of analysis, building ***B2*** is flagged as anomalous (see Figure 15(b)). New techniques are needed that can perform SFPM at multiple analysis scales.

New research is also needed to explore the use of frequent patterns such as CSTPs to drive models that can predict future crime. With the recent interest in predictive policing, pursuing this direction may help in enhancing intervention strategies via effective preparedness [37]

6 Conclusions

This chapter explored Spatial Frequent Pattern Mining (SFPM), which is a data-driven approach to crime pattern analysis. We identified the benefits of data driven approaches in the face of large spatio-temporal crime datasets and highlighted that they are useful in reducing human effort. Hypotheses regarding real world phenomena pertaining to crime can be generated only after analysts have evaluated the results of the SFPM process. Hence, SFPM simply reduces the effort an analyst might have to undertake to formulate a meaningful hypothesis regarding the nature of crime patterns. Pressing research needs, including new SFPM methods for outbreak detection that account for crime distributions along street networks and analysis across multiple scales were identified with specific examples.

Acknowledgement: We are extremely grateful to Dr. Ned Levine and Mr. Ronald E. Wilson for thoughtful exchanges and helpful advise during the collaboration between the University of Minnesota and the National Institute of Justice on the CrimeStat program. We are especially grateful to Mr. Tom Casady, Public Safety Director, Lincoln, NE (then Police Chief) for providing us real crime datasets. We would like to thank Kim Koffolt and the members of the spatial database and spatial data mining group at the University of Minnesota, Twin-Cities. Special thanks, to Dr. Mete Celik and Dr. Betsy George for discussions pertaining to street network hotspots. This research was funded by the U.S. Army grant, W9132V-09-C-0009, USDOD-NGA grants, HM1582-08-1-0017 and HM1582-07-1-2035.

References

[1] Lincoln City Crime Dataset. Lincoln City police department, Lancaster County, NE, USA, 2007.

[2] Arcgis desktop edition 10. Environmental Systems Research Institute(ESRI) Redlands,

CA, USA, 2010.

[3] N. H. T. S. Administration. Data-driven approaches to crime and traffic safety (ddacts): Operational guidelines. http://www.nhtsa.gov/DOT/NHTSA/Traffic% 2520Injury%20Control/Articles/Associated%20Files/811185.pdf, 2009.

[4] L. Anselin. Spatial econometrics. A companion to theoretical econometrics, pages 310-330, 1988.

[5] L. Anselin, J. Cohen, D. Cook, W. Gorr, and G. Tita. Spatial analyses of crime. Criminal Justice 2000: Measurement and analysis of crime and justice, 4, 2000.

[6] L. Anselin, I. Syabri, and Y. Kho. Geoda: An introduction to spatial data analysis. Geographical Analysis, 38(1):5-22, 2006.

[7] D. Beavon, P. Brantingham, and P. Brantingham. The influence of street networks on the patterning of property offenses. Crime prevention studies, 2:115-148, 1994.

[8] C. Beck and C. McCue. Predictive Policing: What Can We Learn From Wal-Mart and Amazon About Fighting Crime in a Recession? The Police Chief, 76(11), 2009.

[9] P. Brantingham and P. Brantingham. Environmental criminology. Sage Publications Beverly Hills, CA, 1981.

[10] P. Brantingham and P. Brantingham. Patterns in crime. Macmillan, 1984.

[11] P. Brantingham and P. Brantingham. Nodes, paths and edges: Considerations on the complexity of crime and the physical environment. Journal of Environmental Psychology, 13(1):3-28, 1993.

[12] P. Brantingham and P. Brantingham. Criminality of place: Crime generators and crime attractors. European Journal on Criminal Policy and Research, 3(3):5-26, 1995.

[13] P. Brantingham, P. Brantingham, and U. Glasser. Computer simulation as a research tool in criminology and criminal justice. Criminal Justice Matters, 58:19-20, 2005.

[14] P. Brantingham, P. Brantingham, M. Vajihollahi, and K. Wuschke. Crime analysis at multiple scales of aggregation: a topological approach. Putting Crime in its Place, pages 87-107, 2009.

[15] P. Brantingham, U. Glasser, P. Jackson, and M. Vajihollahi. Modeling criminal activity in urban landscapes. Mathematical Methods in Counterterrorism, pages 9-31, 2009.

[16] P. Brantingham, U. Glasser, B. Kinney, K. Singh, and M. Vajihollahi. A computational model for simulating spatial aspects of crime in urban environments. In Systems, Man and Cybernetics, 2005 IEEE International Conference on, volume 4, pages 3667-3674.IEEE, 2006.

[17] P. Brantingham, B. Kinney, U. Glasser, P. Jackson, and M. Vajihollahi. Mastermind: Computational modeling and simulation of spatiotemporal aspects of crime in urban environments. Artificial Crime Analysis Systems: Using Computer Simulations and Geographic Information Systems. Information Science Reference, 2008.

[18] M. Buchanan. Sin Cities: The geometry of Crime. New Scientist, 30th April, 2008.

[19] E. Burgess. Juvenile delinquency in a small city. Journal of the American institute of

criminal law and criminology, 6(5):724-728, 1916.

[20] E. Burgess, R. Park, and E. Burgess. The growth of the city: an introduction to a research project. The City, pages 47-62, 1925.

[21] R. Burke. An introduction to criminological theory. No.: ISBN 978-1-84392-407-4, page 422, 2009.

[22] M. Celik, S. Shekhar, B. George, J. Rogers, and J. Shine. Discovering and quantifying mean streets: A summary of results. 2007.

[23] S. Chainey and J. Ratcliffe. GIS and crime mapping, volume 6. John Wiley & Sons Inc, 2005.

[24] H. Chen, W. Chung, J. Xu, G. Wang, Y. Qin, and M. Chau. Crime data mining: a general framework and some examples. Computer, pages 50-56, 2004.

[25] R. Clarke and M. Felson. Routine activity and rational choice. Advances in criminological theory. Transaction Publishers, 2004.

[26] L. Cohen and M. Felson. Social change and crime rate trends: A routine activity approach. American sociological review, pages 588-608, 1979.

[27] J. Eck, S. Chainey, J. Cameron, M. Leitner, and R. Wilson. Mapping crime: Under-standing hot spots. 2005., National Institute of Justice (NIJ), Washington D.C.

[28] U. Glasser and M. Vajihollahi. Computational modeling of criminal activity. Intelligence and Security Informatics, pages 39-50, 2008.

[29] M. Goodchild. The fundamental laws of giscience. Invited talk at University Consortium for Geographic Information Science, University of California, Santa Barbara, 2003.

[30] M. Goodchild. Giscience, geography, form, and process. Annals of the Association

of American Geographers, 94(4):709{714, 2004.

[31] A. Guerry, S. Lacroix, A. Silvestre, and P. Girard. Essai sur la statistique morale de la France. Crochard, 1833.

[32] K. Harries. Crime and the environment. Number 1035. Charles C Thomas Pub Ltd, 1980.

[33] K. Harries. Mapping crime: principle and practice. US Dept. of Justice, Office of Justice Programs, National Institute of Justice, 1999.

[34] A. Hey, S. Tansley, and K. Tolle. The fourth paradigm: data-intensive scientific discovery. Microsoft Research, 2009.

[35] Y. Huang, S. Shekhar, and H. Xiong. Discovering colocation patterns from spatial datasets: A general approach. IEEE Transactions on Knowledge and Data Engineering, 16(12):1472-1485, 2004.

[36] R. Jervis. 'Flash mobs' pose challenge to police tactics. USA Today,
URL: http://www.spotcrime.com/ .

[37] M. M. Kiney. Targeting the next crime. Star Tribune, 26th January, 2011.

[38] J. Kinney, P. Brantingham, K. Wuschke, M. Kirk, and P. Brantingham. Crime attractors, generators and detractors: land use and urban crime opportunities. Built environment, 34(1):62-74, 2008.

[39] J. LeBeau and M. Leitner. Introduction: Progress in research on the geography of crime. The Professional Geographer, 63(2):161-173, 2011.

[40] I. Lee and V. Estivill-Castro. Exploration of massive crime data sets through data mining techniques. Applied Arti_cial Intelligence, 25(5):362-379, 2011.

[41] M. Leitner and J. Kent. Bayesian journey-to-crime modelling of single and multiple

crime-type series in Baltimore County, MD. Journal of Investigative Psychology and Offender Profiling, 6(3):213-236, 2009.

[42] N. Levine. Crime mapping and the crimestat program. Geographical Analysis, 38(1):41-56, 2006.

[43] N. Levine. CrimeStat: a spatial statistics program for the analysis of crime incident locations (v 3.3). Ned Levine & Associates, Houston, TX, and the National Institute of Justice, Washington, DC, 2010.

[44] N. Levine and R. Block. Bayesian journey-to-crime estimation: An improvement in geographic profiling methodology. The Professional Geographer, 63(2):213-229, 2011.

[45] 2007 Nebraska Cornhusker Football Team, Wikipedia 2007. URL: http://en.wikipedia.org/wiki/2007_Nebraska_Cornhuskers_football_team

[46] H. Mayhew. London labour and the London poor. London labour and the London poor. 1861.

[47] C. McCue. Data mining and predictive analysis: intelligence gathering and crime analysis. Butterworth-Heinemann, 2007.

[48] P. Mohan, S. Shekhar, J. Shine, and J. Rogers. Cascading spatio-temporal pattern discovery. IEEE Transactions on Knowledge and Data Engineering, 2011.

[49] P. Mohan, S. Shekhar, J. A. Shine, and J. P. Rogers. Cascading spatio-temporal pattern discovery: A summary of results. In 10th SIAM International Conference on Data Mining, pages 327-338, 2010.

[50] P. Mohan, S. Shekhar, J. A. Shine, J. P. Rogers, and Z. Jiang. A spatial neighborhood graph approach to regional co-location pattern discovery: A summary of results. In 19th ACM SIGSPATIAL International Symposium on Advances in Geographic

Information Systems, ACM-GIS 2011, November 1-4, 2008, Chicago, Illinois, USA, To appear, 2011.

[51] P. Mohan, R. E. Wilson, S. Shekhar, B. George, N. Levine, and M. Celik. Should SDBMS support a join index?: a case study from crimestat. In 16th ACM SIGSPATIAL International Symposium on Advances in Geographic Information Systems, ACM-GIS 2008, November 5-7, 2008, Irvine, California, USA, Proceedings, page 37, 2008.

[52] G. Mohler, M. Short, P. Brantingham, F. Schoenberg, and G. Tita. Self-exciting point process modeling of crime. Journal of the American Statistical Association, 106(493):100-108, 2011.

[53] M. Kullorff. SatScan: User Guide, URL: http://www.satscan.org/techdoc.html .

[54] A. Murray, I. McGu_og, J. Western, and P. Mullins. Exploratory spatial data analysis techniques for examining urban crime. British Journal of Criminology, 41(2):309, 2001.

[55] G. Oatley and B. Ewart. Data mining and crime analysis. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2011.

[56] A. Okabe, K. Okunuki, and S. Shiode. Sanet: a toolbox for spatial analysis on a network. Geographical Analysis, 38(1):57-66, 2006.

[57] M. O'Leary. The mathematics of geographic pro_ling. Journal of Investigative Psychology and Offender Profiling, 6(3):253-265, 2009.

[58] D. Oliver, A. Bannur, J. Kang, S. Shekhar, and R. Bousselaire. A k-main routes approach to spatial network activity summarization: A summary of results. In 2010 IEEE International Conference on Data Mining Workshops, pages 265-272. IEEE Computer Society, 2010.

[59] P. Phillips and I. Lee. Crime analysis through spatial areal aggregated density patterns. GeoInformatica, pages 1-026, 2011.

[60] A. Quetelet. A treatise on man. William and Robert Chambers, Edinburg, U.K., 1842.

[61] Kulldorff M. A spatial scan statistic. Communications in Statistics: Theory and Methods, 1997; 26:1481-1496

[62] J. Ratcliffe. The hotspot matrix: A framework for the spatio-temporal targeting of crime reduction. Police Practice and Research, 5(1):5{23, 2004.

[63] J. Ratcliffe. Crime mapping: Spatial and temporal challenges. Handbook of Quantitative Criminology, pages 5-24, 2010.

[64] D. Rossmo. Geographic profiling. CRC Press, 2000.

[65] D. Rossmo and L. Velarde. Geographic profiling analysis: principles, methods and applications. Crime Mapping Case Studies, pages 33-43, 2008.

[66] S. Shekhar and S. Chawla. Spatial Databases: a tour. Pearson Education, 2003.

[67] S. Shekhar, P. Zhang, and Y. Huang. Spatial data mining. Data Mining and Knowledge Discovery Handbook 2nd Ed., O.Maimon and L.Rokach (eds.) , pages 837-854, 2010, Springer.

[68] M. Short, A. Bertozzi, and P. Brantingham. Nonlinear patterns in urban crime: Hotspots, bifurcations, and suppression. SIAM Journal on Applied Dynamical Systems, 9:462, 2010.

[69] M. Short, P. Brantingham, A. Bertozzi, and G. Tita. Dissipation and displacement of hotspots in reaction-diffusion models of crime. Proceedings of the National Academy of Sciences, 107(9):3961, 2010.

[70] M.S.Scott and K. Dedel. Assaults in and around bars(2nd edition). Problem Oriented Guides for Police, Problem Specific Guides, 1:1-78, 2006.

[71] R. Stone. Investigations of excess environmental risks around putative sources: statistical problems and a proposed test. Statistics in Medicine, 7(6):649{660, 1988.

[72] H. Tasdoven and B. Sahin. Crime data mining as a decision making tool. International Journal of Public Policy, 6(3):278{287, 2010.

[73] N. Tate and P. Atkinson. Modelling scale in geographical information science. John Wiley & Sons Inc, 2001.

[74] L. Wasserman. All of statistics: a concise course in statistical inference. Springer texts in statistics. Springer, 2004.

[75] I. Jung, M. Kulldorff, O.J. Richard. A spatial scan statistic for multinomial data. Statistics in Medicine, 2010, 29(18), 1910-1918. John Wiley and Sons.

[76] M. Leitner, M. Barnett and J. Kent and T.Barnett. The Impact of Hurricane Katrina on Reported Crimes in Louisiana: A Spatial and Temporal Analysis. The Professional Geographer, 2011, 63(2), 244-261. Taylor and Francis.

[77] T. Nakaya , K. Yano. Visualising crime clusters in a space-time cube: An exploratory data-analysis approach using space-time kernel density estimation and scan statistics. Transactions in GIS, 14:223-239, 2010.

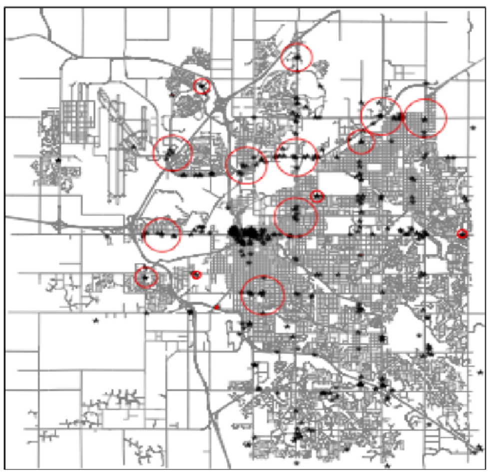
[78] J.L. LeBeau. Demonstrating the analytical utility of GIS for police operations: A final report. National Criminal Justice Reference Service, 2000

[79] E.S. Jefferis. A multi-method exploration of crime hot spots: SaTScan results. National Institute of Justice, Crime Mapping Research Center, 1998.

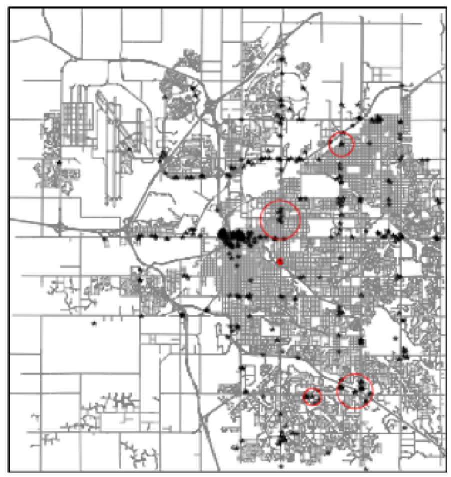
[80] M. Leitner and J. Kent (2009) Bayesian Journey to Crime Modeling of Single- and Multiple Crime Type Series in Baltimore County, MD. In Levine, N. and R. Block (eds.) Bayesian Journey-to-Crime Estimation. Special issue of Journal of Investigative Psychology & Offender Profiling, 6 (3): 213-236.

Appendix

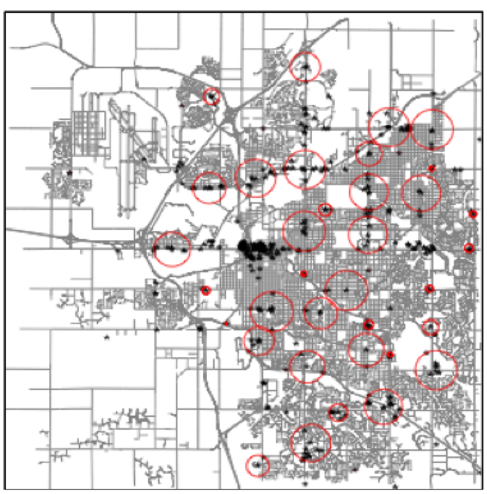
I Spatial crime outbreak (additional types)



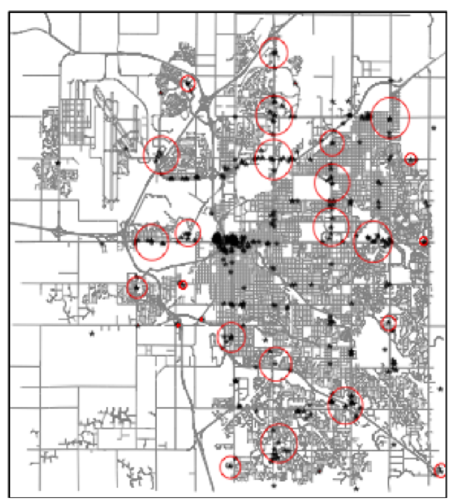
(a): Auto-theft outbreaks



(b): Traffic offense outbreaks



(c) Burglary Outbreaks



(d) Larceny Outbreaks

Figure A.1: Significant crime type outbreaks

Figure A.1 shows additional crime type outbreaks in the vicinity of bars in Lincoln, NE during the year 2007.

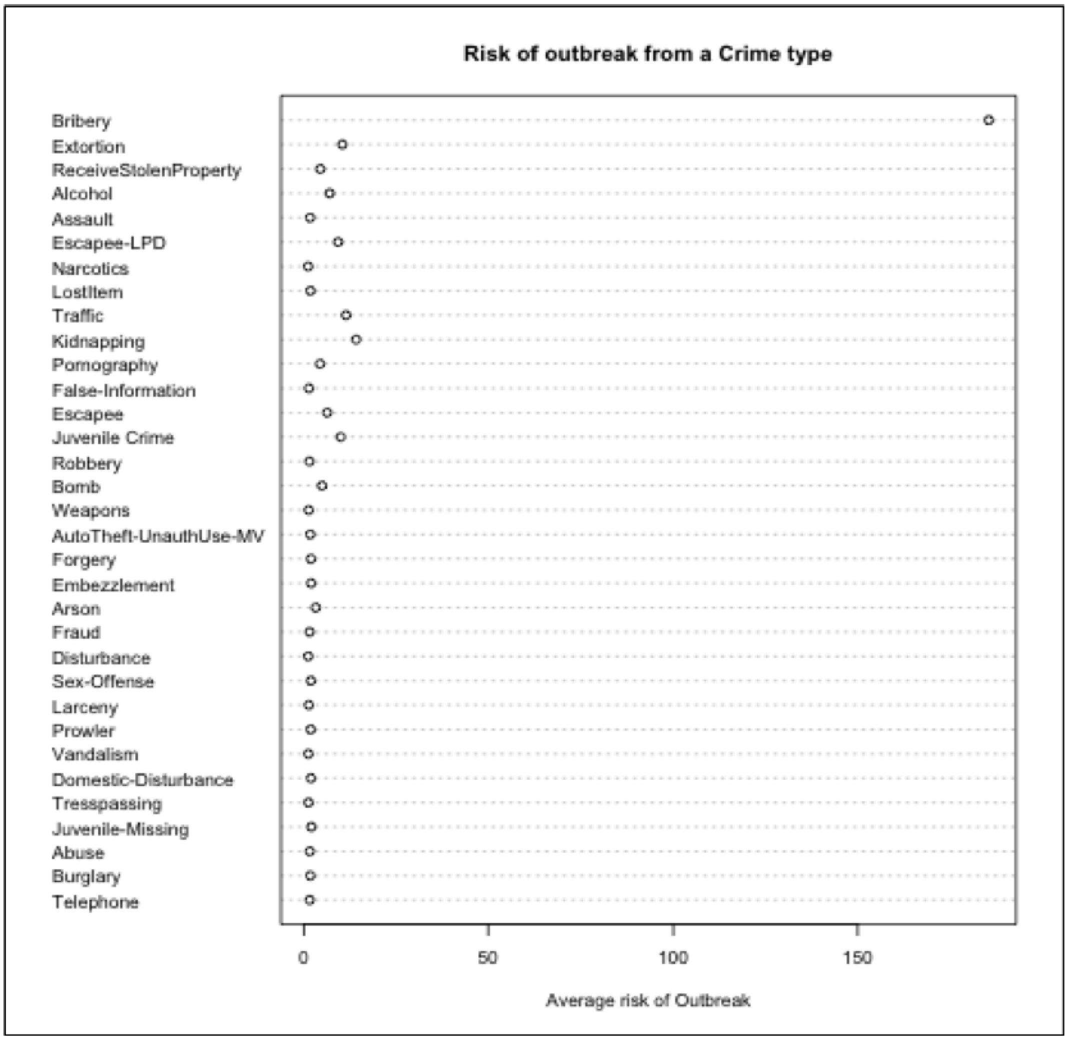


Figure A.2: Average Risk of a crime type being within any outbreak

II Space-time crime outbreak

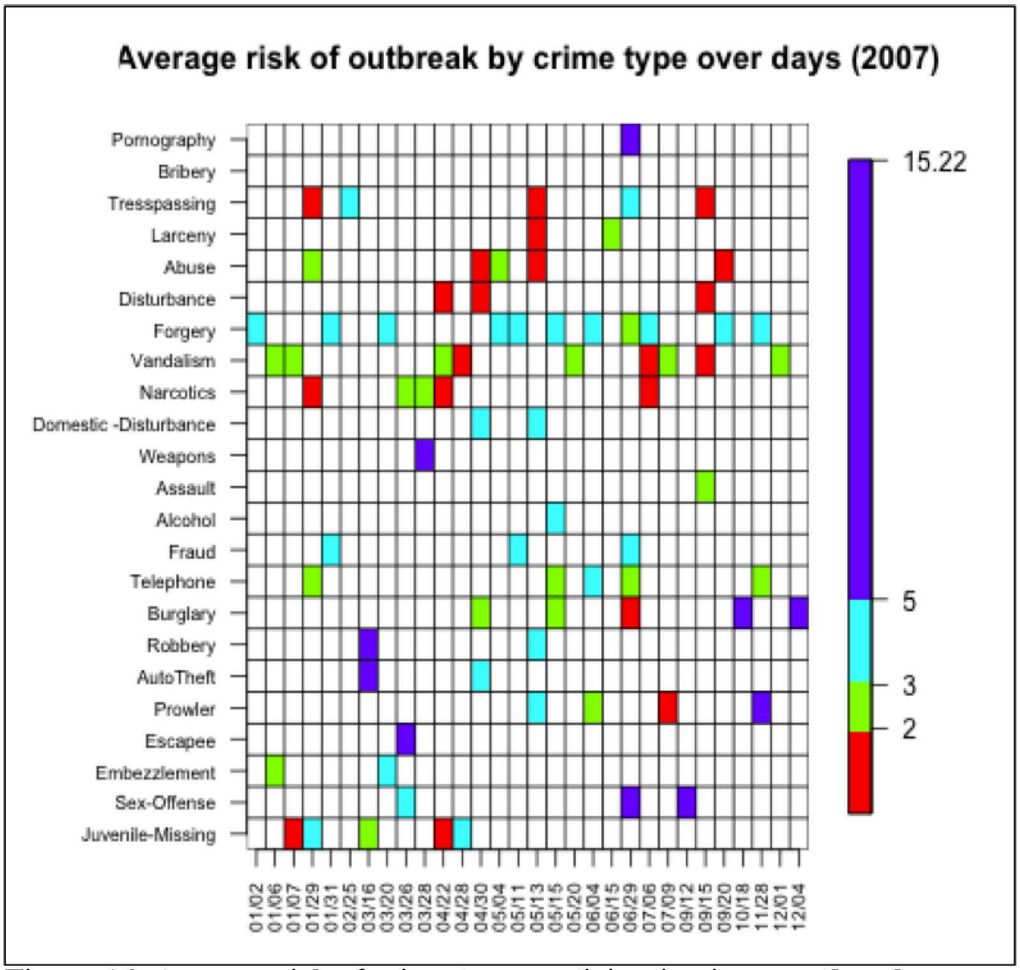


Figure 16: Average risk of crime type participating in an outbreak (Best viewed in color, Log2(Average Risk) is shown)